结合LHD算法与MLP神经网络优化铝合金清洗剂配方

林瑞祥[1]　　何培浩[1]
[1]（达高工业技术研究院（广州）有限公司，广东省 广州市　510700）

摘要：
针对传统的配方设计方法繁琐费时、物料损耗多、优化效果差的缺陷，将拉丁超立方设计（LHD）算法与多层感知器（MLP）神经网络相结合，形成铝合金清洗剂配方优化的新方法。利用LHD算法全面探索参数空间和MLP神经网络的高效建模能力，成功优化了铝合金清洗剂配方。经过优化的配方，净洗力从87.9 %，提升到98.24 %，腐蚀量从4.2 mg，下降到0.3 mg。对模型预测数据与实验数据进行回归直线拟合后，大于0.98，证明预测与实验数据保持一致性。

关键词：铝合金清洗剂；拉丁超立方设计（LHD）算法；多层感知器（MLP）神经网络
分类号：TP393

The optimization of aluminum alloy cleaner formulas by combining the LHD algorithm with MLP neural networks.

Ruixiang Lin[1]　Peihao He[1]
[1](Dagao Industrial Technology Research Institute (Guangzhou) Co., Ltd., Guangzhou 510700, China)

Abstract：
Addressing the drawbacks of conventional formula design methods, such as being cumbersome, time-consuming, resulting in material wastage, and yielding poor optimization outcomes, a novel approach for optimizing the formula of aluminum alloy cleaners is proposed by combining the Latin Hypercube Design (LHD) algorithm with the Multilayer Perceptron (MLP) neural network. Leveraging the comprehensive exploration of parameter space using the LHD algorithm and the efficient modeling capability of the MLP neural network, the formula for aluminum alloy cleaners was successfully optimized. With the optimized formula, the cleaning efficiency increased from 87.9% to 98.24%, and the corrosion rate decreased from 4.2 mg to 0.3 mg. Regression analysis of the model's predicted data against experimental data yielded a correlation coefficient greater than 0.98, demonstrating consistency between predicted and experimental data.

Keywords：　Aluminum alloy cleaner; Latin Hypercube Design (LHD) algorithm; Multilayer Perceptron (MLP) neural network

1　引言

铝合金在现代工业中扮演着重要角色，其轻质、高强度和良好的耐腐蚀量质使铝合金成为航空航天、汽车制造、建筑等领域重要的常用材料。然而，在加工和生产过程中，铝合金常受到各种外部污染物的侵害，这直接影响了其表面质量

和性能。因此，高效铝合金表面清洗剂成为当前工业界急需解决的挑战之一。

传统的清洗剂配方设计通常基于大量试验和经验积累，这种方式既费时又昂贵。近年来，计算机辅助方法和机器学习技术在复杂试验空间中设计实验得到了广泛关注[1]。特别是，机器学习方法，如神经网络、遗传算法等，具有在复杂参数空间中进行高效优化的能力，极大地减少了传统试验方法所需的时间和成本[1]。传统的响应面曲线和正交实验往往受限于特定参数范围的探索，无法全面覆盖复杂的参数空间[2]，这可能限制了对最优配方的发现。

本文应用拉丁超立方设计（Latin Hyper Design，LHD）算法[3]结合神经网络（Neural Network）方法呈现出更为突出的优势[4]。LHD 算法以更为高效和均匀的方式在整个参数空间中进行采样，这使得它能够实现更为全面的探索。同时，神经网络作为一种强大的机器学习工具，具备对于复杂数据的高效学习和建模能力[3]。与传统的方法相比，神经网络能更有效地处理数据集，捕捉到参数之间更为复杂的非线性关系。其极高的准确性使得在铝合金有机碱清洗剂配方优化的过程中，神经网络能够更为精准地预测不同配方的性能，从而为最终的优化结果提供更可靠的依据[8]。这种结合为优化过程提供了更全面、更准确的数据探索和性能预测手段，有助于在清洁剂配方设计领域取得更大的突破和进步。

因此，我们选择将 LHD 算法与多层感知器（MLP）神经网络相结合应用于铝合金有机碱清洗剂配方的优化，是为了充分利用 LHD 算法全面探索参数空间的优势，以及 MLP 神经网络高效建模和预测的特点，以加速最优配方的发现并实现高效、环保的清洁效果。这种结合将为工业界提供更具创新性和可持续性的解决方案，推动清洁生产和可持续制造的发展。

2 实验

本文工作遵循图 1 所示的流程图，主要分为三个核心部分：首先是训练集数据的收集，接着是 MLP 神经网络模型的训练，最后是 LHD 算法的数据生成以及寻找最佳清洗分数，进而逆向推导出相应的输入配方。尽管这个流程所需的实验量相比正交实验等传统实验更大，但它能够涵盖更为复杂的参数空间，从理论上讲，能够寻找到性能最佳的配方。此外，训练出的 MLP 神经网络模型还可用于过程中的数据监控，用以追踪净洗力和腐蚀量。并且本文的所有代码均由 Python 语言

编写，因此非常容易迁移到其他平台或体系中进行训练，为该方法的推广应用提供了便利条件。

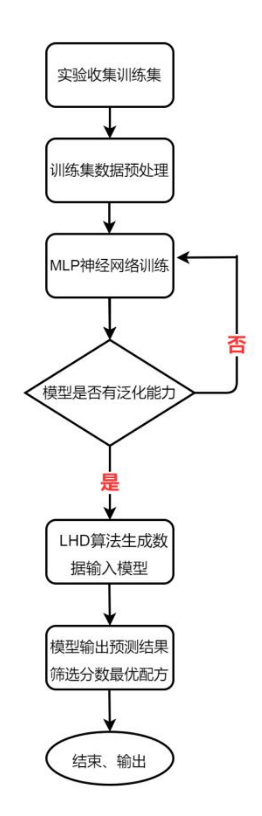


图 1　LHD 算法结合 MLP 神经网络优化清洗剂配方流程图

Figure 1　The LHD algorithm combined with the MLP neural network optimizes the flow chart of the cleaning agent formulation

2.1 准备 MLP 神经网络训练集

训练集数据由我们进行实验收集。训练集中输入数据一共有五个维度，分别为清洗剂基础配方包含的原料[6,7,11]：单乙醇胺、酒石酸钾钠、去离子水、镁铝缓蚀剂 DX5819、自制非离子型表面活性剂。训练集中输出数据一共有两个维度，分别为基材的净洗力和腐蚀量。

基材选择为 2A12 铝合金，常应用在航空航天、汽车制造等领域。净洗力和腐蚀量分别按照 GB/T35759-2017 中 5.7 和 5.8 规定的工艺进行操作和结果计算。通过实验的训练集一共有 150 组数据，基本满足神经网络训练最小所需数据量。

2.2　训练集预处理

由于训练集比较小，所以在进行训练之前，需要用 Z-Score 方法对训练集进行清洗，除去异常值。Z-Score 方法可以帮助确定一个数据点在数据集中的相对位置，并用标准差的倍数来表示数据点与均值之间的距离。Z-Score 的计算方式

如下：

$$Z = \frac{X - \mu}{\sigma} \tag{1.1}$$

式（1.1）中，X 是数据点的值；μ 是数据集的均值；σ 是数据集的标准差。Z-Score 的绝对值越大，表示数据点与均值的偏差程度越大。常见的做法是将 Z-Score 大于某个阈值的数据点定义为异常值。本文将阈值设置为 2，超过 2 个标准差的值都会从训练集中被删除。经过处理后，训练集的数量从 150 组下降到 143 组。

由于配方系数总和为 1，所以无需进行归一化处理。但数据量较少，在 MLP 神经网络训练之前，会对训练集使用数据增强函数，利用正态分布来生成噪声，并将这些噪声添加到每个样本中，以创建新的增强样本，减少过拟合的风险。

2.2 MLP 神经网络模型构建

为了进一步增强模型的泛化能力，我们选择了在 TensorFlow 2.15 框架中使用 Sequential 模型[9]来构建 MLP 神经网络模型，该模型结构如图 2 所示。在这一模型中，我们通过配置全连接层，以及设置一系列隐藏层和输出层的方式，使得模型能够更有效地学习和表示特征。具体而言，我们的输入包括五个特征，这五个特征代表着清洗剂配方中的五种原料。与此同时，我们的输出层由两个节点组成，分别用于净洗力和腐蚀量的预测。

通过全连接层的设计，我们能够充分利用各层神经元之间的连接关系，使得模型能够更好地理解和捕捉数据中的特征和规律。这样的结构设置使得神经网络模型能够在配方优化过程中更精确地预测所需的性能指标，从而为配方设计和性能预测提供了可靠的基础。

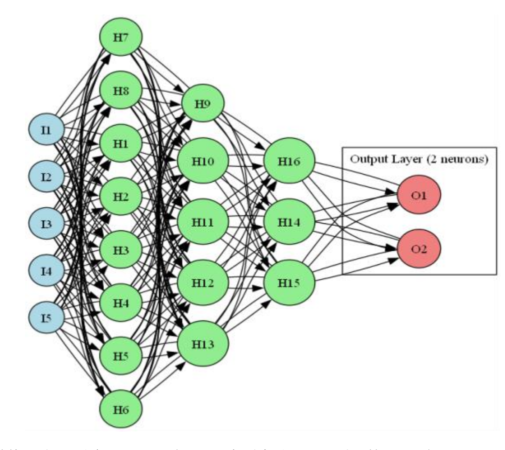


图 2　MLP 神经网络模型，输入层有五个特征，隐藏层有三层，输出层有两个节点。

Figure 2 In the MLP neural network model, there are five features in the input layer, three layers in the hidden layer, and two nodes in the output layer

我们采用了一个包含三个隐藏层的神经网络架构。这些隐藏层分别设定为 8、5 和 3 个神经元。在每个隐藏层之后，我们使用了 ReLU（Rectified Linear Unit）激活函数进行激活，并添加了 Dropout 层以及引入了 L2 正则化（权重衰减）机制，以限制模型的复杂性和减少过拟合的风险。

考虑到净洗力和腐蚀量数据的特征，我们的输出层通过 Sigmoid 激活函数进行激活，确保输出数据的值在 0 到 1 之间。这样的设定有助于确保输出数据在合理范围内，并使其符合问题的特性。优化模型的过程中，我们采用了 Adam 优化器，并使用了学习率衰减策略，以更好地平衡模型的收敛速度和优化效果[10]。在模型训练过程中，我们还引入了早停策略，减少过拟合风险。同时还通过 K 折交叉验证，我们对模型进行了多次训练和评估，并且使用了损失函数、准确率和 F1 分数来监控模型的性能，以达到良好的泛化能力[5]。

2.4 MLP 神经网络模型评估

我们设置了 10 组数据作为测试集，分别对模型预测的净洗力和腐蚀量，通过回归拟合曲线进行对比，进一步评估模型的泛化能力。

2.5 LHD 算法生成数据输入 MLP 神经网络模型

LHD 算法能够在多维空间内均匀分布样本点，以确保输入数据在各个维度上都有足够的覆盖，避免了随机抽样可能导致的不均匀性，提供更全面的数据集，并且不需要任何初值输入。具体操作中，由于配方的自由度为 4，所以利用 LHD 算法对除了去离子水以外的 4 种原料进行随机抽样。得益于 LHD 算法的高效性，可以很轻松生成十多万条配方作为输入数据。这些数据输入到 MLP 神经网络模型中，能够快速得到净洗力和腐蚀量。

因为净洗力要求越大越好，腐蚀量要求越小越好，所以我们定义一个无量纲的清洗分数，公式如下：

$$\text{清洗分数}=\omega_1 * \frac{1}{\text{净洗力}} + \omega_2 * \text{腐蚀量} \quad (1.2)$$

式（1.2）中，ω_1和ω_2都是权重系数，在本文中，为了平衡数量级，设置ω_1=50，ω_2=1，代表对净洗力和腐蚀量关注程度是一致的。我们可以

通过设置权重系数，调整对净洗力和腐蚀量的关注度。清洗分数还可以根据研究对象进行调整，如可以增加成本要素。

由式（1.2）可知，最小的清洗分数对应的配方就是在当前权重系数下最优的配方。

3 结果与讨论

3.1 MLP 神经网络模型训练评估

在 MLP 神经网络模型的训练评估中，我们观察到随着训练的 epoch 增加，图 3 中显示的损失函数（Loss）在训练集和验证集上持续下降，而准确率（Accuracy）持续上升直至 1。这趋势表明模型的训练过程顺利进行。另外，在表 1 中我们记录了 5 次 K 折交叉验证的结果，包括每次折叠的损失函数、准确率和 F1 分数。这些数据显示出在不同验证集上的性能差异不大，表明了模型具有较好的泛化性能。这些结果表明，随着训练次数的增加，模型在训练集和验证集上的性能稳步提升，而交叉验证的结果也显示了模型的稳健性。

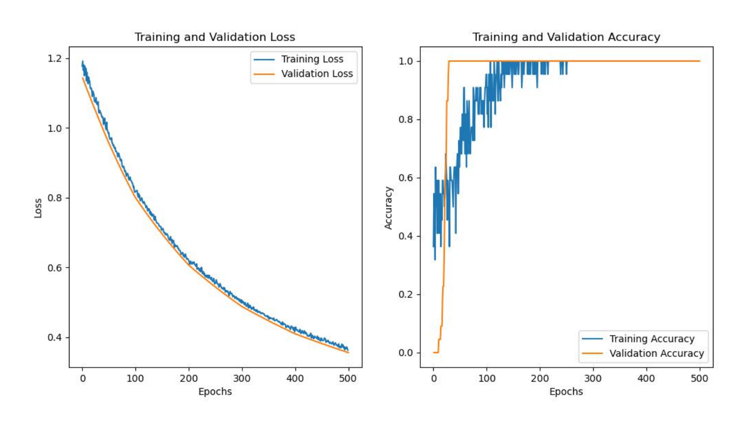


图 3　MLP 神经网络模型训练过程中损失函数（Loss）和准确率（Accuracy）的趋势。

Figure 3　The trend of loss function and accuracy during the training process of MLP neural network model.

表 1　在 5 次 K 折交叉验证中，每次折叠的训练损失函数、训练准确率和 F1 分数。

Table 1　Training loss function, training accuracy, and F1 score for each fold in 5 K-fold cross-validations.

折叠次数	Training Loss	Training Accuracy	F1 score
第 1 次折叠	0.3555	1.000	0.9633
第 2 次折叠	0.3776	1.000	0.9742
第 3 次折叠	0.3555	1.000	1.000

第 4 次折叠	0.3487	1.000	0.9479
第 5 次折叠	0.3564	1.000	1.000

为了全面评估模型的泛化性能，我们准备了 10 组未曾用于训练的数据作为测试集。通过图 4 和图 5 的观察，我们发现尽管 MLP 神经网络模型对净洗力和腐蚀量的预测与实验数据之间存在一定程度的差异，但通过对预测结果和实验结果进行回归直线 y = x 拟合，发现净洗力和腐蚀量的回归直线 R^2 值均超过 0.98。这一结果表明预测值与实验值之间具有高度的一致性，可接受程度较高。综合考虑这些结果，我们可以得出结论：基于所采集的训练集，MLP 神经网络模型表现出良好的准确性和泛化能力。

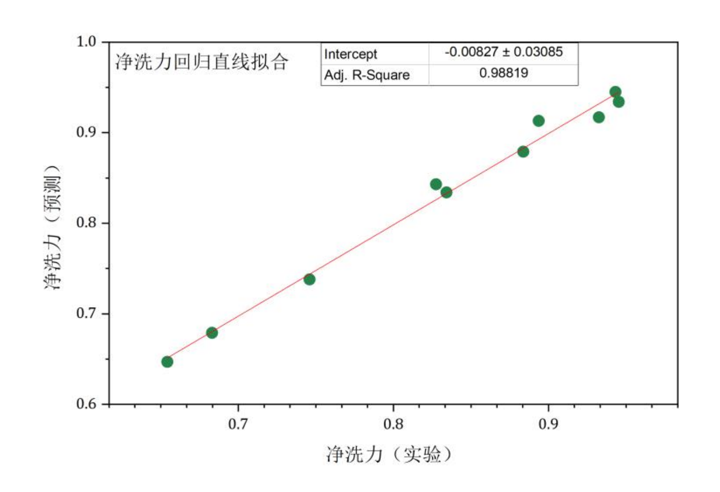


图 4　MLP 神经网络模型预测的净洗力与实验数据的回归拟合曲线

Figure 4　The regression fitting curve of the net washing power predicted by the MLP neural network model and the experimental data

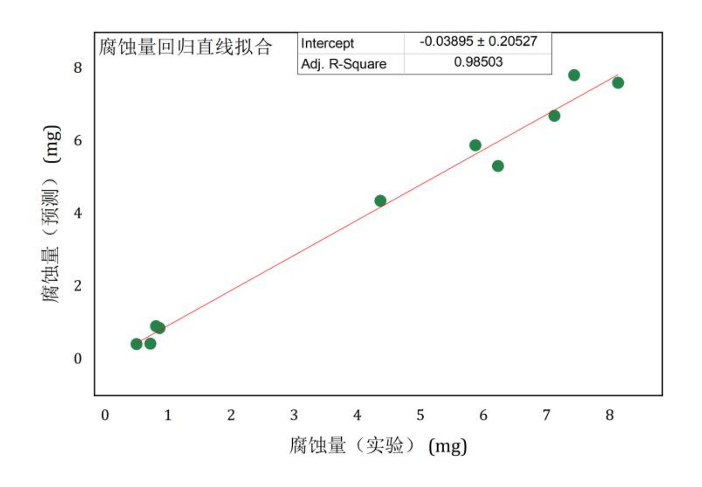


图 5　MLP 神经网络模型预测的腐蚀量与实验数据的回归拟合曲线

Figure 5　The regression fitting curve of the corrosion amount predicted by the MLP neural network model and the experimental data

3.1　优化结果对比

经过验证 MLP 神经网络模型的泛化性能，我们引入了拉丁超立方设计（LHD）算法，并合理设定了数据范围，分别生成了不同规模的数据集，包括 1 万、10 万和 100 万组作为输入。将这些数据集输入 MLP 神经网络模型后，迅速获得了预测的净洗力和腐蚀量，并根据设定的权重系数快速输出了最小清洗分数及其对应的输入数据。

从 100 万组数据输入集的优化结果来看，发现清洗分数从实验最佳配方中的 61.08 下降到 51.20，净洗力从 87.9 %提升到最高的 98.24 %，同时腐蚀量也从 4.2 mg 降至最低的 0.31 mg。这些显著改善了清洗剂的性能。

进一步对比 1 万、10 万和 100 万组数据的优化配方结果，呈现出随着数据集规模增加，优化结果逐步改善的趋势。然而，与 100 万组数据相比，10 万组数据的优化配方性能提升幅度较小，暗示着数据集规模超过 10 万后，边际效应显著。鉴于 LHD 算法和 MLP 神经网络的高效性，建议在使用过程中充分利用这些特点，尽可能地增加数据集规模，以达到预期的最佳优化效果。

表 2　实验最佳配方，1 万、10 万和 100 万组数据分别优化的预测配方，以及对应的净洗力、腐蚀量、清洗分数

Table 2 The optimal formula for the experiment, the predicted formula optimized by 10,000, 100,000 and 1 million sets of data, as well as the corresponding net cleaning power, corrosion amount, and cleaning score

	单乙醇胺	无机盐	去离子水	表面活性剂	缓蚀剂	净洗力	腐蚀量	清洗分数
实验最佳配方	15.00 %	0.50 %	82.30 %	1.50 %	0.70 %	87.90 %	4.20 mg	61.08
1 万组数据优化预测配方	19.83 %	2.63 %	74.36 %	1.44 %	1.74 %	94.78 %	1.08 mg	53.75
10 万组数据优化预测配方	18.37 %	1.94 %	75.46 %	2.31 %	1.92 %	97.93 %	0.77 mg	51.82
100 万组数据优化预测配方	24.73 %	2.27 %	69.42 %	2.15 %	2.43 %	98.24 %	0.31 mg	51.20

为了保证优化结果的可靠性，我们对优化后的配方进行三次重复性实验，结果取平均值，同时对预测结果和实验结果进行回归直线 y = x 拟合，并计算R^2值。从表 3 中可以看出，R^2值都接近 1，说明实验结果与预测结果保持一致性。

表 3 对 1 万、10 万和 100 万组数据分别优化的预测配方进行实验重复，得到净洗力、腐蚀量和清洗分数的实验值，以及预测值和实验值的R^2值

Table 3 The experimental replicates of the predicted formulas optimized for 10,000, 100,000 and 1 million sets of data were carried out to obtain the experimental values of net washing power, corrosion amount and cleaning fraction, as well as the predicted value and experimental value

	1 万组数据优化预测配方	10 万组数据优化预测配方	100 万组数据优化预测配方	R^2
预测净洗力	94.78 %	97.93 %	98.24 %	0.9967
实验净洗力	94.40 %	98.00 %	98.10 %	
预测腐蚀量	1.08 mg	0.77 mg	0.31 mg	0.9851
实验腐蚀量	0.81 mg	0.64 mg	0.23 mg	
预测清洗分数	53.75	51.82	51.20	0.9939
实验清洗分数	53.87	51.65	51.20	

4 结论

本文通过结合 LHD 算法在整个参数空间中进行均匀采样和 MLP 神经网络高效建模的特性，成功实现了铝合金清洗剂配方的高效优化。经过此方法优化后的配方，性能相比传统实验结果有明显提高，净洗力从 87.9 %，提升到 98.24 %，腐蚀量从 4.2 mg，下降到 0.3 mg。并且优化效率高，为工业界提供了更创新、更可持续的解决方案。同时由于 MLP 神经网络模型预测准确率高，可以非常方便地调整清洗分数的权重，或者增加关注因素，达到最佳的预期效果。此方法不仅可以在铝合金清洗剂配方开发领域发挥作用，还可以在其他领域的配方开发以及使用过程中性能监控发挥重要作用。

参考文献：

[1] 尤杨. 复杂试验空间下试验设计方法及应用研究[D].国防科技大学, 2022.

You Yang. Method and Application of Experimental Design in Complex Test Space[D].National University of Defense Technology, 2022.

[2] Xiaobing Shang, Tao Chao, Ping Ma, et al. An efficient local search-based genetic algorithm for constructing optimal Latin hypercube design[J]. Engineering Optimization, 2020, Volume 6 , Issue 1.

[3] Joseph, V. R. & Hung, Y. Orthogonal-maximin Latin hypercube designs. [J]. Stat. Sin, 2008, 18, 171–186.

[4] LeCun, Y., Bengio, Y., Hinton, G. Deep learning. [J].Nature, 2015, 521, 436–444.

[5] Sengul Mert Y. , Song Yao, Van Duin Adri C. T. INDEEDopt: a deep learning-based ReaxFF parameterization framework. [J]. npj Computational Materials. 2021, Volume 7 , Issue 1.

[6] 刘宁华, 郭艳红, 赖心翘, 等. 7 系铝合金的绿色环保型清洗剂的开发[J]. 电镀与涂饰, 2023, 42 (14): 37-41.

LIU N H, GUO Y H, LAI X Q, et al. Development of an environmentally friendly cleaning agent for 7-series aluminum alloys [J]. Electroplating & Finishing, 2023, 42 (14): 37-41.

[7] 邓迎诚, 徐新颖, 王飞, 等. 自制碱性化学清洗剂对换热器用 3 种金属腐蚀行为的影响[J]. 电镀与涂饰, 2023, 42 (12): 17-22.

DENG Y C, XU X Y, WANG F, et al. Corrosion behaviors of three metals used in heat exchangers after being treated with an alkaline cleaner [J]. Electroplating & Finishing, 2023, 42 (12): 17-22.

[8] HAN W, WANG Y, SU X , et al. A multi-objective optimization problem research for amphibious operational mission of shipboard helicopters[J]. Chinese Journal of Aeronautics, 2023, 36(09):256-279.

[9] Joseph, V. R., Dasgupta, T., Tuo, R., Wu, C. F. [J]. Sequential exploration of complex surfaces using minimum energy designs. Technometrics, 2015, 57, 64–74.

[10] Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980. 2015.

[11] 季伟, 孟繁东, 潘科宇等. 铝合金表面油污清洗剂的制备及其腐蚀量研究[J]. 电镀与精饰, 2021, 43(02):38-43.

JI W, MENG F D, PAN K Y, et al. Preparation and Corrosion Research of a Water-Based Cleaning Agent for Aluminum Alloy [J]. Plating and Finishing, 2021,43(02):38-43.

(通讯作者: 林瑞祥 E-mail:20162432007@m.scnu.edu.cn)

作者贡献声明：

林瑞祥：提出研究思路，设计研究方案，代码撰写，数据处理和分析，论文撰写；

何培浩：设计研究方案，进行实验。